Subjective Comparison of Consumer Television Technologies for 3D Visualization

Jesús Gutiérrez, Fernando Jaureguizar, and Narciso García

Abstract—**This paper presents a comparison among different consumer 3D display technologies by means of a subjective assessment test. Therefore, four 55-in displays have been considered: one autostereoscopic display, one stereoscopic with polarized passive glasses, and two with active shutter glasses. In addition, a high-quality 3D video database has been used to show diverse material with both views in high definition. To carry out the test, standard recommendations have been followed considering also some modifications looking for a test environment more similar to real home viewing conditions, with the objective of obtaining more representative conclusions. Moreover, several perceptual factors have been considered to study the performance of the displays, such as picture quality, depth perception, and visual discomfort. The obtained results show interesting issues, like the performance improvement of active shutter glasses technology, the high performance of the polarized glasses technology in terms of quality and comfort, and the need of improvement of the autostereoscopic displays to complement the visual comfort to reach a global high-quality visual experience.**

Index Terms—**Depth perception, quality of experience, subjective evaluation, 3D displays, visual discomfort.**

I. Introduction

NOWADAYS, 3D video technology is widely spread across the consumer market providing the users the possibility of watching 3D content not only at cinemas, but also at households with 3D television sets, and even in mobile devices, such as smartphones.

However, there is still the need to research and develop 3D video systems to achieve the total acceptation of the consumers, which has not been gained during these years due to many issues. Especially, because of the lack of high quality 3D video content and due to various factors of the visualization technologies, for instance the need of using specific glasses to watch 3D videos with some visualization systems, which do not provide a totally satisfactory Quality of Experience (QoE) to the viewers.

In this process to reach an essential improvement in 3D video technology, it is crucial to know the opinions and reception of the people regarding 3D video systems and applications, since they are finally addressed to the end users. Therefore, there is a strong active work on the field of users' QoE evaluation, in order to provide a robust basis on the development of 3D video technology, as it happened in other fields, like traditional video and audio coding.

Moreover, several new perceptual factors are involved in the visual experience related to 3D video in comparison with traditional video, such as depth perception and visual discomfort [1]. These factors play a crucial role in the viewers' QoE, thus a reliable evaluation is essential to know the perceptual response of people to the 3D video systems and applications under study.

In addition to the human factors involved in the 3D QoE, there are system factors related to the processes of capture, coding, transmission, and visualization of 3D content [2]. In fact, all the steps involved in the processing chain of 3D video could affect the visual experience of the end users. This is the reason why several research works have been proposed in the literature analyzing aspects like capturing effects [3], coding schemes [4], transmission events [5], and visualization issues [6]. However, more studies about the quality perceived by the users of 3D video systems are needed, dealing with new approaches and technologies to achieve a performance improvement and to understand the users' expectations, especially in relation to the capabilities of consumer 3D visualization technologies. In fact, only a few works have been published regarding these issues, which are extremely important as it has been shown in similar fields, like in conventional television sets [7] or mobile devices [8].

Although the most practical way to assess the QoE perceived by the observers would be by objective metrics that automatically provide estimations of what people perceive, their reliability is not satisfactory in many cases, especially in 3D video, where the research on objective metrics is an ongoing work [9]. Therefore, it is inevitable to turn to subjective assessment tests to obtain robust conclusions about the users' QoE.

Thus, taking this into account, this paper presents a subjective assessment test evaluating and comparing different consumer 3D display technologies, in order to understand people's acceptation, and to analyze possible factors and trends that could help in the improvement of 3D video development. With this aim, several perceptual factors have been evaluated (e.g., picture quality, depth quality, visual discomfort), and standard recommendations for subjective tests with 3D video have been considered, as well as realistic viewing conditions. In addition, a wide high-quality and freely available 3D video database has been used, in order to obtain representative and reliable results.

The rest of the paper is structured as follows. In Section II a detailed description of the technologies used for 3D visualization is given. In Section III, a review of the state-of-the-art in relation with subjective assessment tests with 3D displays is presented. Section IV presents the details of the subjective assessment test carried out for the present work, whose results are shown and explained in Section V. Finally, Section VI provides some general conclusions.

II. 3D Display Technologies

Since the presentation of the first stereoscope by Charles Wheatstone in 1838 [10], a great development on technologies for visualizing 3D content has taken place. Especially important is the progress reached in the last years thanks to decisive technological and hardware advances. This development, which is still ongoing [11], has allowed the apparition of various alternatives to watch 3D content, even in the consumer market.

A possible classification of the displays is based on their degree of parallax, having two-view, horizontal-parallax, and full-parallax displays [12]. The first is the simplest type of 3D displays, which provides one view for the left eye and one for the right eye of the observers to allow stereoscopic perception. On the other hand, horizontal-parallax displays provide simultaneously multiple views of the same scene, so the observers could see different perspectives changing horizontally their position within some defined regions. Finally, full-parallax displays provide variations in what is being perceived by the observers with both horizontal and vertical head movements.

As the scope of this work is focused on consumer displays for 3D video visualization for home entertainment, the three most common technologies have been considered: stereoscopic displays with passive polarized glasses and active shutter glasses, and multiview autostereoscopic displays [13]. However, the interesting properties and capabilities of other visualization technologies, such as head-mounted displays, holographic displays, or integral imaging systems should not be forgotten [11], [12].

A. Stereoscopic Displays With Passive Polarized Glasses

These two-view systems present simultaneously the left and right views to the observer using spatial multiplexing, so both stereo images are interleaved line-by-line. Obviously, this entails a reduction of the vertical resolution per eye to the half.

To watch these displays correctly, passive polarized glasses are required to make possible that each eye sees the corresponding stereo view, which is emitted by the display through a filter film located in the screen that polarizes the horizontal lines with alternate polarization.

B. Stereoscopic Displays With Active Shutter Glasses

These are also two-view systems, however, in this case the two views are presented to the viewers alternatively, using time multiplexing. Thus, it is necessary to use active shutter glasses (blocking alternately the light to each eye), synchronized with the TV (usually by infrared systems), to see each time the corresponding view in each eye. Although the spatial resolution of the images is conserved, it is necessary to, at least, double the framerate of the displays to reduce annoying flickering effects.

C. Multiview Autostereoscopic Displays

This type of displays are horizontal-parallax systems that do not require any kind of glasses to watch 3D content, since each view is sent to different spatial regions in front of the display. Thus, when the observer is located at the correct position, each eye receives the corresponding view. The main technologies used are lenticular lenses and parallax barriers. The first type is based on the use of adjacent vertical lenses located in front of the display limiting the visibility of each column of pixels to a specific zone. Therefore, the resolution of the display is divided by the number of zones (i.e., views). On the other hand, parallax barrier displays use vertical slits located in front of the screen, creating contiguous opaque and translucent regions and generating a set of viewing zones where the observers could experience stereoscopic perception. Also, in this type of displays, the resolution is divided by the number of views [12].

III. Related Work

Subjective tests are the most reliable way to study the effects of 3D visualization on the visual experience of the users. Therefore, several subjective experiments have been carried out analyzing different factors related to watching 3D content. Especially important are those studies concerning visual discomfort or fatigue [6], commonly experienced by the viewers of 3D video and caused by the use of glasses, motion artifacts [14], and vergence-accommodation conflicts [15]. Depth perception is other factor affecting 3D video visualization that is being deeply investigated [16], as well as distortions introduced by 3D visualization technologies, like crosstalk between views [17], [18].

Another important issue regarding 3D visualization is to subjectively evaluate the performance of the different technologies that are available to watch 3D content, and compare different approaches to identify their advantages and drawbacks, to determine the best use cases for each solution. For example, Slanina *et al.* [19] compared the performance of different consumer 3D display technologies: a 32″ stereoscopic LCD with passive glasses, a 42″ stereoscopic plasma with active shutter glasses, and an 80″ projection system with shutter glasses. Various factors influencing the users' QoE (e.g., depth perception, image sharpness, visual comfort) were evaluated by means of a non-standard methodology designed by the authors and using side-by-side videos from a proprietary database. The results showed comparable performance among the considered systems regarding depth perception, but differences in image sharpness and visual comfort. Other example is the study carried out by Kaller *et al.* [20] comparing the angular characteristics of a polarization multiplex stereoscopic display and a time multiplex display, concluding that the time multiplex technology is better in terms of image quality and perceived depth. Also, Yun *et al.* [18] carried out a comparison between common stereoscopic displays with active and passive glasses in terms of perceptual resolution and crosstalk, finding out interesting results, such as the lower crosstalk in passive displays. Another interesting study was presented by Rerabek and Ebrahimi [21] comparing three different 3D portable displays that are commercially available, in terms of depth quality and overall quality provided by restitution techniques based on stereo and motion parallax. The results show that the motion parallax method offers a better overall quality, while stereo parallax provides better depth quality.

Taking this into account, the work presented in this paper aims at complementing these previous studies by offering a deep comparison among the three most common 3D display technologies addressed to the consumer market: stereoscopic displays with polarized glasses, with active shutter glasses, and multiview autostereoscopic displays. To avoid the possible influence of the screen size on the comparison, displays of the same size where used. In addition, a standard evaluation methodology was used, although the test environment was set mimicking typical living room viewing conditions to obtain more representative results, since these displays are usually oriented to home entertainment. Furthermore, videos with Full High-Definition (Full-HD) for ach view were used, selected from a high-quality video database available for the community, especially developed for 3D video subjective studies [22].

IV. Subjective Assessment

A. Evaluation Methodology

The use of an appropriate methodology is crucial to obtain reliable results in the subjective tests. Therefore, the recommendation ITU-R BT.2021 [23] was followed, as it provides the guidelines to carry out subjective tests of 3DTV systems, extending the traditional standard ITU-R BT.500 [24] for subjective assessments with conventional television.

Since the goal of this work was to compare various display technologies (coding or transmission artifacts do not degrade the test sequences) and a simultaneous comparison is not feasible (e.g., due to the use of different glasses), a single stimulus method was selected. Therefore, all the test sequences were evaluated by observers in each display at a time, presenting a different random order of the displays in each test session.

In particular, for each test sequence, the observers were asked to evaluate the picture quality, depth quality, and visual discomfort, which are the primary perceptual dimensions in 3D video [23]. A five-grade quality scale was used to evaluate picture quality and depth quality, while a five-grade comfort scale was used for visual discomfort, as defined in [23].

After the evaluation of all the test sequences in a display, the observers were asked to provide a global score for the primary perceptual factors for that specific display, in addition to global evaluations for the additional perceptual dimensions: naturalness and sense of presence [23]. Also, an overall evaluation of the whole visual experience utilizing the quality scale was requested. Moreover, after the evaluation of all the displays, the observers were asked to rank them considering their global QoE.

The test sessions started with an exhaustive explanation of the experiment (focusing attention on the definitions of the perceptual dimensions extracted from the recommendation ITU-R BT.2021 [23]), a visual screening of the observers, and the visualization of training sequences in all the displays, in order to familiarize the observers with the different TVs and provide a reference for the subsequent evaluations. After evaluating each display, the observers could have a break time while the following display was set up. The total duration of the whole assessment session was around 50 minutes.

To collect the opinions from the observers, questionnaires with boxes were used, where they were asked to write a mark for the corresponding evaluation. All the questionnaires were available in Spanish and English, for foreigners.

TABLE I
Characteristics of the Displays

ID	Model	Technology	Luminance (cd/m^2)	Release Year
SO	Sony KDL-55HX920	Stereoscopic: Shutter glasses TDG-BR250/B	250 (100)	2011
LG	LG 55LW980S	Stereoscopic: Polarized glasses AG-F315	360 (140)	2011
SA	Samsung UE55HU8500L	Stereoscopic: Shutter glasses SSG-5100GB	260 (80)	2014
TO	Toshiba 55ZL2G	Autostereoscopic: Glasses-free	290	2012

B. Environment and Equipment

Four consumer displays were selected to carry out the tests, all of them having a screen size of 55″, in order to avoid the influence of other factors than the technology in the visual experience. In particular, the displays used in the experiment are detailed in Table I, where the values of luminance were measured without glasses and through them (in parenthesis). The complete technical specifications of the displays can be found in [25]–[28].

The displays SO and LG are typical HD LED consumer displays, while SA is a curved UHD LED display. According to manufacturers, this type of screen improves some aspects in comparison with typical flat screens, such as immersiveness, sense of depth, contrast, sharpness, field of view, and effective viewing angle. On the other hand, some drawbacks have been observed, like worse effects of onscreen reflections, geometry issues when watching from the sides or a limited viewing area around the center of the screen to appreciate the advantages of the curved display. Some of these aspects were taken into account in the analysis of the results, except from those related to viewing positions, since they were not considered in the realization of the tests, where the observers were seated around the center of the screen.

The autostereoscopic display uses the lenticular technology to provide nine views (taking a stereo pair as input and using a proprietary algorithm to create seven more perspectives) with an effective resolution of 720p, thanks to a total panel resolution of 3840×2160 pixels and a pixel layout to group each nine pixels. The display has an integrated camera that detects the observers' faces to adjust the viewing zones. In particular, it recognizes the person closest to the center of the TV and sets four more adjacent viewing zones. This process is especially important to assure the correct position of the observers within one of the viewing zones, which are approximately 50 cm wide, and allow the observers to properly perceive the nine views without blurring effects that appear when they are not well positioned.

To playback the test sequences a PC with a graphic card GeForce GTX 560 and Nvidia 3D Vision drivers were used to run the Stereoscopic Player [29], providing the output to the

TABLE II
SEQUENCES USED IN THE TESTS

ID	V1	V2	V3	V4	V5	V6	V7	V8	V9	T1	T2
Sequence	Barrier gate	Basket	Hall	Lab	Phone call	Soccer	Tree branches	Umbrella	Big Buck Bunny	News report	Boxers
Description	Outdoor scene of a car crossing a barrier gate	Basket training	Indoor scene of people in a hall	Lab assistants working	Person talking by phone in an office	Soccer scene	Scene of tree leaves and branches	Person playing with an umbrella	Animation movie	News report imitation	Boxing training
Duration	16s	16s	16s	16s	16s	16s	16s	13s	16s	16s	16s
SI/TI	59/21	71/41	82/5	53/12	36/13	89/38	101/14	74/19	60/30	53/4	50/19
DSI/DTI	20.4/15.4	11.7/9.7	17/7	17.8/10	21.6/11.8	24.7/18.1	23/13.6	17/15.2	26.1/35.2	21.6/8.7	24.4/18
D+/D-	6/9	-14/26	7/-3	6/22	17/15	7/10	3/9	5/17	21/3	16/6	9/3

displays in frame packing format (Full-HD each stereo view). 1080p/24 Hz. However, due to severe flickering effects with the SO display, this TV was used in frame packing format 720p/60 Hz.

Since the displays used are addressed to consumer market, the test conditions should be kept as far as possible similar to a typical real-life usage. Thus, the displays were configured in standard mode with default settings (e.g., medium brightness, maximum contrast, etc.) and disabling any image processing algorithm that they could implement (e.g., increase depth, edge enhancement, etc.), apart from the generation of the nine views of the autostereoscopic display and the upscaling of SA and SO displays to fit the whole screen.

Furthermore, one of the main aspects in the design of a subjective test is the selection of the appropriate test environment. Although commonly the international recommendations are used to establish a standardized test environment, there is a recent trend focused on carrying out subjective tests under realistic conditions, similar to those the users have at their homes [5]. The objective of these approaches is to obtain more representative results from the experiments, because using standardized environments could distance the observers from real viewing conditions. Therefore, since the systems under study in the present work are displays addressed to the consumer market to be used in a domestic context, the test environment was set mimicking home conditions, as shown in Fig. 1. Thus, a couch was placed at around 2.1 meters from the display position (which corresponds to a viewing distance of 3H and it is within the viewing distance ranges of all the displays used), and the ambient lightning conditions were controlled to avoid disturbing reflections.

Fig. 1. Test environment.

C. Test Material

A total of nine stereoscopic Full-HD (1920×1080 each view) videos were used as source material for the test. The sequences were selected in order to cover a wide range of contents and spatial, temporal, and depth characteristics. The selected sequences are described in Table II, where their main properties can be found. In particular, *SI* and *TI* are the spatial and temporal perceptual indicators, as defined in ITU-T P.910 [30], based on the standard deviation over the pixels of the Sobel-filtered frames, and on the motion difference between adjacent frames, respectively. *DSI* and *DTI* are the spatial and temporal perceptual indicators computed similarly over the depth maps, and $D+$ and $D-$ are the maximum crossed and uncrossed disparities. All the sequences were selected from the freely available database published in [22], where more details can be found, except "Big Buck Bunny" that is a segment of an open source movie [31]. The sequences V1-V9 were used for the test sequence generation, while the sequences T1 and T2 were used in the training process of the observers explained in Section IV-A. The original sequences were converted using *Ffmpeg* [32] to *avi* files containing uncompressed video with 24 fps without audio, to avoid the introduction of possible coding degradations and synchronization errors, and the influence of audio in the observers' QoE.

Since one of the objectives of the experiment was to compare the QoE perceived by the observers with stereoscopic and monoscopic content, the videos were also displayed in 2D using twice the left view as inputs, so the viewing conditions were the same than in 3D (frame packing format, display properties, observers wearing glasses, etc.). To equalize the conditions in which the videos were evaluated, the test sequences were generated concatenating the 18 videos (9 in 3D and 9 in 2D), using different randomizations for each observer and display, with the condition of not watching the same source consecutively. After each clip of the test sequence, a message in a grey background was displayed during ten seconds providing the time to the observers to write their scores in the questionnaire, and indicating

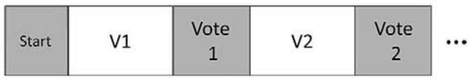

Fig. 2. Scheme of the test sequences.

the number of the corresponding box to write them, as depicted in Fig. 2.

D. Observers

A total of 26 observers (9 females, 17 males) participated in the tests, all of them having normal or corrected vision as reported in a previous visual screening carried out before each test session, in which a Randot test were used for 3D vision. The ages of the participants were ranged between 20 and 50, with an average age of 31. The observers were rewarded for their participation in the tests, and a maximum of two observers were allowed in each test session. Furthermore, a screening of the results scores provided by the observers was carried out according to the method recommended in ITU-R BT.500 [24], but no outliers were detected.

V. Results

In the following subsections the main results obtained from the experiment are exposed. The results shown in the figures were obtained computing the Mean Opinion Scores (MOS) from the evaluations provided by the observers in the questionnaires. In addition, the 95% confidence intervals are represented according to the computation recommended by the standard ITU-R BT.500 [24].

A. Picture Quality and Depth Quality

The results obtained for the evaluation of picture quality are shown in Fig. 3 for the 3D (in columns) and 2D (lines with markers) sequences. Regarding 3D picture quality, the general trend is that SA and LG displays provide better image performance, while the autostereoscopic display (TO) is the weakest system in this aspect, due to the loss of sharpness caused by the reduction of the resolution by the number of views. An important point that could be extracted from these results is the difficulty in perceiving the resolution loss of the stereoscopic display with polarized glasses (LG). Also, the effects of visualizing the videos in 720 p format in the SO display instead of 1080p are not easily discernible, since the picture quality of this display is similar to the other stereoscopic displays. Considering the results for the monoscopic test videos there is more similarity among the performances of each display.

To study the statistical significance of the differences among MOS values of picture quality for different displays and contents, analyses of variance (ANOVA) have been performed. In particular, a two-way repeated measures ANOVA showed significant effects of the display ($p < 0.0001$ for 3D content and $p < 0.05$ for 2D), the content ($p < 0.0001$ for 3D and 2D), and a significant interaction between both factors ($p < 0.0001$ for 3D and 2D). Moreover, one-way repeated measures analyses were applied to study the simple main effects, revealing a significant effect of content for each display ($p < 0.0001$ for all cases in 2D and 3D) and a significant effect of display for each content ($p < 0.05$ at least), except for V1, V5 and V9 for 3D, and V1,

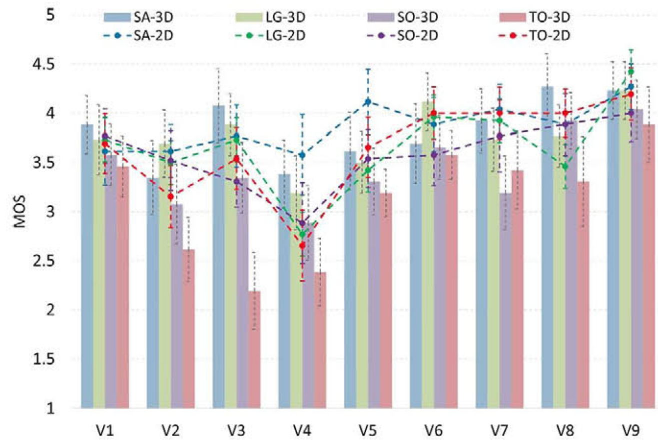

Fig. 3. Picture quality.

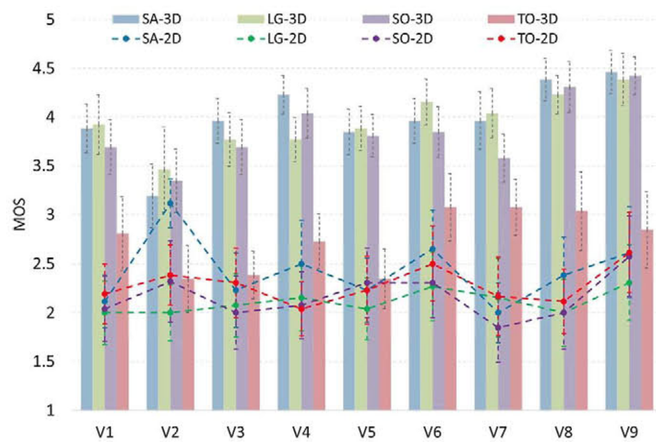

Fig. 4. Depth quality.

V3, V7 and V9 for 2D. For the 3D versions, this could be caused by the lower spatial detail of V1 and V5 in comparison with the rest of the clips, could lessen the impact of resolution losses of LG and TO displays, and the typically better visualization of computer graphics movies in 3D for V9. On the other hand, the less significant influence of the display in the 2D contents is in accordance with the high performances of the displays working in a conventional mode and without resolution losses. It is also worth noting that, in general, the poorest results were obtained for V2 and V4. In fact, V2 is a scene of a basketball match with a wide camera panning that could introduce some motion artifacts in 3D affecting the perceived quality. In the case of V4, poor results were obtained both in 2D and 3D, which could be caused by a strong lightning change contained in the sequence.

In relation with depth quality (see Fig. 4) similar results have been obtained for the three stereoscopic displays (SA, LG, and SO) providing a good performance for all 3D contents. However, the results confirm the worse ability of autostereoscopic displays to provide depth sensation to the .4.5 observers. Again, V2 is the content with lower scores for 3D depth quality, which, as aforementioned, could be caused by motion artifacts produced by the camera panning. On the other hand, V8 and V9 provide the best results, since the first sequence is a scene recorded from a close distance and taking care on avoiding the frame effect, in which the umbrella offers a good sensation as it is going out of the screen, and V9 is a cartoon video in

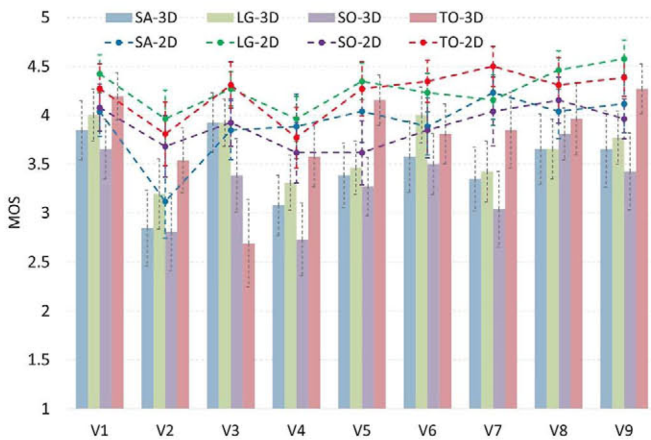

Fig. 5. Visual discomfort.

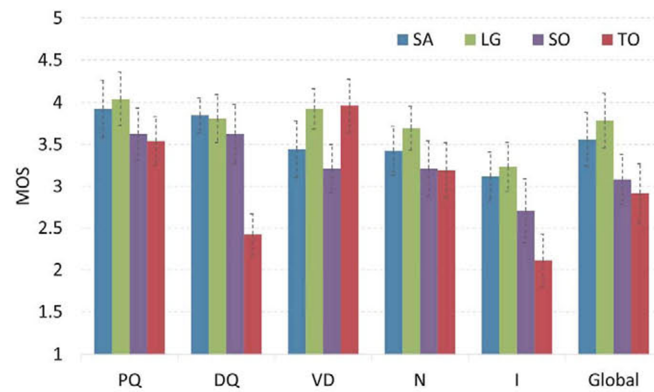

Fig. 6. Global evaluations.

which depth issues are easily controlled. Finally, as expected, the results show a poor depth quality of the test videos when displayed in 2D for all displays.

Similarly as for picture quality, ANOVA analyses have been performed with the MOS values obtained for depth quality. The two-way repeated measures ANOVA also showed a significant effect of the display and the content ($p < 0.0001$ for 3D and 2D for both effects), and a significant interaction between both factors ($p < 0.005$ for 3D and 2D). In addition, one-way repeated measures analyses revealed a significant effect of content for each display (at least $p < 0.05$ for all cases in 2D and 3D), except for the LG in 2D. Also, a significant effect of display was proved for all contents in 3D ($p < 0.0001$ for all cases), but only for V2, V4 and V8 in 2D at $p < 0.05$ level. Again, the less significant influence of the display in 2D reflects a uniform performance of the TVs in monoscopic mode, while LG display offers a more stable response regarding the depth sensation of content visualized in 2D. In opposition, the good results of SA obtained in some monoscopic videos, like in V2 based on a camera panning in a basketball match, could reflect an improvement in depth sensation thanks to the curved screen. However, this is not evidenced with 3D videos as exhibited by the similarity of the results for the three stereoscopic displays.

B. Visual Discomfort

The results obtained from the evaluations of visual discomfort are shown in Fig. 5. In general, the best performance in terms of visual comfort for 3D content is provided by the autostereoscopic display. On one hand, this fact could be related to the reduction of the depth sensation offered by the autostereoscopic display, since it can lessen the impact of excessive disparity (as happens in V7) or frame effects (as appear in V4, where the disparity concordance between the content and the screen frame is violated). On the other hand, apart from the annoyance caused to the visual system, some observers considered the use of glasses a source of discomfort in their viewing experience, thus this could be another reason to the better performance of the TO display in this aspect. It is worth noting that for some contents the autostereoscopic display is significantly better than the stereoscopic technology with active shutter glasses, possibly due to the flickering effects that can be perceived with these systems in some scenes with high motion activity (V2 has the greatest TI) or bright colors (V5).

Regarding the visual comfort in 2D test videos, in general, better results were obtained in comparison with the corresponding 3D versions of the sequences, and similar performances are provided by all displays. Generally better results were reported by the autostereoscopic display, since no glasses are needed, and the stereoscopic with polarized glasses, caused by the absence of flickering effects.

Again, the two-way repeated measures ANOVA carried out with the MOS values obtained for visual discomfort reflected a significant effect of the display and the content ($p < 0.0001$ for 3D and 2D for both effects), and a significant interaction between both factors ($p < 0.0001$ for 3D and $p < 0.01$ for 2D). Furthermore, one-way repeated measures analyses revealed a significant effect of content for each display (at least $p < 0.005$ for all cases in 2D and 3D), and a significant effect of display for all contents (at least $p < 0.05$ for all cases) except for V8 in 3D, and V4, V7 and V8 in 2D. In the case of V8, it could be caused by the good properties of the content (slow motion, no frame effects, no excessive disparity, etc.), while for V4 and V7 the visual comfort with the monoscopic versions is considerably better than in 3D, probably due to frame effects and excessive disparity, respectively. A similar behavior could be observed with V2 that could be produced by the appearance of motion artifacts in 3D.

C. Global Display Comparison

As mentioned in Section IV-A, after the evaluation of all the test videos in one display, the observers were asked to provide global evaluations of that display in terms of the five primary and additional perceptual dimensions [23], and a score for the overall QoE. From the results collected in the questionnaires, the MOS values were computed and they are depicted in Fig. 6. Concerning picture quality (PQ) and depth quality (DQ), the results confirm the conclusions described in Section V-A regarding the better performance of SA and LG in terms of picture quality, and the poor ability of the autostereoscopic display to provide depth sensation.

Looking into visual discomfort (VD), the best results are provided by the TO and LG displays, mainly thanks to the glasses-free technology in the first case and the absence of flickering

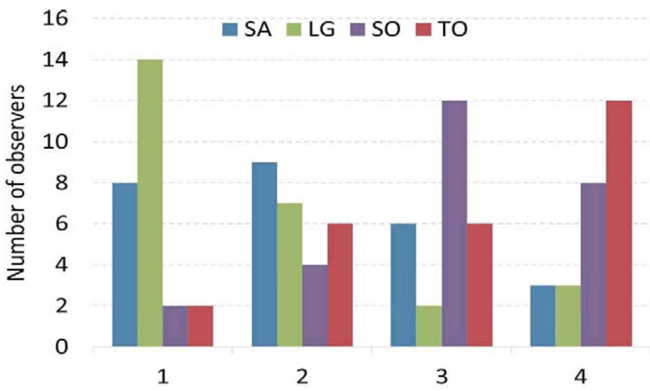

Fig. 7. Observers displays ranking.

effects in the latter. The impact of flickering effects that appear using active shutter glasses are also reflected in the significantly worse results that the SO display obtains in this perceptual dimension.

In relation with the additional perceptual dimensions, for naturalness (N) the LG display seems to offer a better perceptual realism, while in the case of sense of presence (I), which is one of the aspects of the viewers' QoE that could be enhanced with curved displays (such as SA), similar results were obtained among the stereoscopic displays, while the autostereoscopic display offers the worst immersiveness.

Taking into account the evaluations of the overall QoE provided by each display, it can be observed that the SA and LG displays are the best options for 3D visualization in a consumer environment. On one hand, this shows the good performance of stereoscopic displays with polarized glasses, in spite of the loss of resolution due to the spatial multiplex of the stereo views. These displays provide a high picture quality, good depth sensation, and the glasses are generally more comfortable than active shutter glasses and do not cause flickering effects. On the other hand, the evolution of the technology of stereoscopic displays with active shutter glasses, reducing the discomfort of the glasses, is evidenced, since the SA display provides one of the best performances. This fact, in addition to the increase of screen resolution offering a better picture quality, improves considerably the QoE provided by the older SO display. Finally, the results also reflect the low capabilities of the autostereoscopic displays to provide a satisfactory 3D sensation, in spite of the advantage of not needing glasses in terms of comfort.

To analyze the statistical significance of the differences among the MOS values obtained for each perceptual dimension and each display, a one-way repeated measures ANOVA was applied, showing a significant effect of the display in all dimensions (PQ: $p < 0.05$, DQ: $p < 0.0001$, VD: $p < 0.0001$, N: $p < 0.05$, I: $p < 0.0001$) and in the global QoE evaluation ($p < 0.0001$).

At the end of the test, the observers were asked to rank the displays according to their preferences. The results for the rankings concerning the overall QoE are shown in Fig. 7. As it can be seen, the favorite alternative was the stereoscopic display with polarized glasses, selected as the best display by more than half the observers, followed by the newest display with active shutter glasses. On the other hand, the autostereoscopic display seems to be the worst option. These results confirm those aforementioned and depicted in Fig. 6.

D. Other Observations

After the evaluation of each display, the observers were allowed to write free-form comments about their visual experience. Mainly, these observations were focused on the discomfort caused by the active shutter glasses (due to flickering effects, reflections or lack of comfort, especially when the observer also wears progressive glasses), changes in the colors in the case of the SO display, and blurring effects in the autostereoscopic display. In addition, the observers were asked whether they felt visual fatigue, and 54% of the observers reported symptoms like eye fatigue and slight dizziness or headache.

VI. Conclusion

Despite the spread of 3D video technologies in the consumer market, there is still the need of research to improve their performances and to satisfy the users' expectations. Thus, to know in depth the perceptual differences among consumer 3D displays, this paper presents a subjective test comparing the three most common technologies for 3D visualization: stereoscopic displays with polarized glasses and with active shutter glasses, and multiview autostereoscopic displays.

The subjective test was designed following the standard recommendations for 3D video evaluation, but the test environment was set trying to mimic real living room conditions to obtain more representative results for consumer displays. Also, several factors influencing the users' QoE were evaluated using a diverse high-quality video database.

Some interesting conclusions could be extracted from the results. Firstly, the stereoscopic display technology, which was the favorite alternative for more than half the observers, offers a notable performance, even though the resolution is reduced due to the spatial multiplex of both stereo views. Secondly, there has been a great improvement of the technology of stereoscopic displays with active shutter glasses, especially reducing discomforting aspects of the glasses, like flickering effects. Finally, autostereoscopic displays still provide a poor 3D visual experience, despite the advantage of not needing glasses to watch 3D content.

Acknowledgment

The authors would like to thank E. Madejón for her help and support on the development of the subjective tests.

References

[1] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet, "New requirements of subjective video quality assessment methodologies for 3DTV," in *Proc. Int. Workshop Video Process. Quality Metrics*, Scottsdale, AZ, USA, Jan. 2010.

[2] P. Le Callet, S. Möller, and A. Perkis, Eds., "Qualinet white paper on definitions of quality of experience," Eur. Network on QoE in Multimedia Syst. Services (COST Action IC 1003), Ver. 1.1, Jun. 3, 2012.

[3] L. Goldmann, F. De Simone, and T. Ebrahimi, "A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video," in *Proc. SPIE-IS&T Electron. Imaging 3D Image Process. Appl.*, Jan. 2010, vol. 7526, Art no 75260S.

[4] K. Wang, M. Barkowsky, R. Cousseau, K. Brunnström, R. Olsson, P. Le Callet, and M. Sjöström, "Subjective evaluation of HDTV stereoscopic videos in IPTV scenarios using absolute category rating," in *Proc. SPIE-IS&T Stereosc. Displays and Appl.*, Feb. 2011, vol. 7863, Art. ID 78631T.

[5] J. Gutiérrez, P. Pérez, F. Jaureguizar, J. Cabrera, and N. García, "Subjective assessment of the impact of transmission errors in 3DTV compared to HDTV," in *IEEE 3DTV Conf.*, Antalya, Turkey, May 2011, pp. 1–4.

[6] M. Lambooij, W. IJsselsteijn, M. Fortuin, and I. Heynderickx, "Visual discomfort and visual fatigue of stereoscopic displays: A review," *J. Imag. Sci. Technol.*, vol. 53, no. 3, pp. 1–14, May 2009.

[7] S. Tourancheau, P. Le Callet, and D. Barba, "Image and video quality assessment using LCD: Comparisons with CRT conditions," *IEICE Trans. Fundam. Electron., Commun. Comput. Sci.*, vol. E91-A, no. 6, pp. 1383–1391, Jun. 2008.

[8] A. Catellier, M. Pinson, W. Ingram, and A. Webster, "Impact of mobile devices and usage location on perceived multimedia quality," in *Proc. Int. Workshop Quality of Multimedia Experience*, Yarra Velley, Australia, Jul. 2012, pp. 39–44.

[9] A. Moorthy and A. Bovik, "A survey on 3D quality of experience and 3D quality assessment," in *Proc. SPIE-IS&T Electron. Imag.*, Burlingame, CA, USA, Mar. 2013, vol. 8651, p. 86510M-1–11.

[10] C. Wheatstone, "Contributions to the physiology of vision I: On some remarkable and hitherto unobserved phenomena of vision," *Philos. Trans. Royal Soc. London*, vol. 18, no. 13, pp. 371–395, Jun. 1838.

[11] J.-Y. Son, B. Javidi, S. Yano, and K.-H. Choi, "Recent developments in 3-D imaging technologies," *J. Display Technol.*, vol. 6, no. 10, pp. 394–403, Oct. 2010.

[12] N. Holliman, N. Dodgson, G. Favalora, and L. Pockett, "Three-dimensional displays: A review and applications analysis," *IEEE Trans. Broadcasting*, vol. 57, no. 2, pp. 362–371, Jun. 2011.

[13] J. M. Cubero, J. Gutiérrez, P. Pérez, E. Estalayo, J. Cabrera, F. Jaureguizar, and N. García, "Providing 3D video services: The challenge from 2D to 3DTV Quality of Experience," *Bell Labs Tech. J.*, vol. 16, no. 4, pp. 115–134, Mar. 2012.

[14] J. Kuze and K. Ukai, "Subjective evaluation of visual fatigue caused by motion images," *Displays*, vol. 29, no. 2, pp. 159–166, Mar. 2008.

[15] M. Emoto, T. Niida, and F. Okano, "Repeated vergence adaptation causes the decline of visual functions in watching stereoscopic television," *J. Display Technol.*, vol. 1, no. 2, pp. 328–340, Dec. 2005.

[16] P. Lebreton, A. Raake, M. Barkowsky, and P. Le Callet, "Evaluating depth perception of 3D stereoscopic videos," *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 6, pp. 710–720, Oct. 2012.

[17] Video Quality Experts Group, "Test plan for crosstalk influences on user Quality of Experience," 2010.

[18] J. D. Yun, Y. Kwak, and S. Yang, "Evaluation of perceptual resolution and crosstalk in stereoscopic displays," *J. Display Technol.*, vol. 9, no. 2, pp. 106–111, Feb. 2013.

[19] M. Slanina, T. Kratochvíl, V. Říčný, L. Boleček, O. Kaller, and L. Polák, "Testing QoE in different 3D HDTV technologies," *Radioeng.*, vol. 21, no. 1, pp. 445–454, Apr. 2012.

[20] O. Kaller, L. Bolecek, and T. Kratochvil, "Subjective evaluation and measurement of angular characteristics of the 3D stereoscopic displays," in *Proc. Int. Conf. Radioelektronika*, Brno, Czech Republic, Apr. 2012, pp. 1–4.

[21] M. Rerabek and T. Ebrahimi, "Comparison of 3D portable display restitution techniques based on stereo and motion parallax," in *Proc. Int. Workshop on Quality of Multimedia Experience*, Yarra Valley, Australia, Jul. 2012, pp. 80–85.

[22] M. Urvoy *et al.*, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in *Proc. Int. Workshop on Quality of Multimedia Experience*, Australia, Jul. 2012, pp. 109–114.

[23] ITU-R, "Subjective methods for the assessment of stereoscopic 3DTV systems," Recommendation ITU-R BT.2021, Aug. 2012.

[24] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Rec. ITU-R BT. 500-13, Jan. 2012.

[25] Sony, "Sony KDL-55HX920 display technical specifications," [Online]. Available: http://www.sony.co.uk/support/en/content/cnt-specs/KDL-55HX920/list

[26] LG, "LG 55LW980S display technical specifications," [Online]. Available: http://www.lg.com/es/television/lg-55LW980s-3d

[27] Samsung, "Samsung UE55HU8500L display technical specifications," [Online]. Available: http://www.samsung.com/es/consumer/tv-av/tv/uhd/UE55HU8500LXXC

[28] Toshiba, "Toshiba 55ZL2G display technical specifications," [Online]. Available: available at: http://www.toshiba.eu/discontinued-products/55zl2/

[29] "Stereoscopic Player," [Online]. Available: http://www.3dtv.at/

[30] ITU-T "Subjective video quality assessment methods for multimedia applications," Recommendation P.910, Apr. 2008.

[31] Blender Foundation, "Big Buck Bunny Project Homepage," [Online]. Available: http://www.bigbuckbunny.org

[32] "Ffmpeg Project," [Online]. Available: https://www.ffmpeg.org/

Jesús Gutiérrez received the Telecommunication Engineering degree (five-year engineering program) from the Universidad Politécnica de Valencia (UPV), Valencia, Spain, in 2008, and the Communications Technologies and Systems Master degree (two-year M.S. program) from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 2011.

Since 2010, he has been a Member of the Image Processing Group at the UPM.

Fernando Jaureguizar received the Telecommunication Engineering degree (6-year engineering program) and Ph.D. degree in telecommunication from the Universidad Politécnica de Madrid (UPM), in 1987 and 1994, respectively.

Since 1987, he has been a member of the Image Processing Group of the UPM, where he is currently an Associate Professor of Signal Theory and Communications.

Narciso García received the Ingeniero de Telecomunicación degree (five-year engineering program) and Ph.D. degree in communications from the Universidad Politécnica de Madrid (UPM), Madrid, Spain, in 1976 and 1983, respectively.

Since 1977, he has been a Member of the faculty of the UPM where he is currently a Professor of Signal Theory and Communications. He leads the Image Processing Group of the UPM.